Mukesh Kumar Sahu
Data Engineer

Data Engineer with 2+ years of hands-on experience in the analysis and development of data science products, developing end-to-end ETL processes, modelling and architecting data pipelines, and intermediate datasets used for downstream Proficient in Big-data, data wrangling, and data processing algorithms, as well as analysing, including Apache Spark, Apache Airflow, Python, SQL, and Pytest. Capable of creating, developing, testing, and deploying highly adaptive, diverse services to translate business and functional qualifications into substantial deliverables.

EDUCATION

M.Tech.
Data Science and Engineering
National Institute of Technology,
Silchar, Assam
2023 – 2025

B.Tech.
Electrical and Electronics Engineering
Silicon Institute of Technology,
Bhubaneshwar, Odisha
2016 – 2020, secured 7.82 CGPA

12th, Science
S.C.S Junior College, Puri, Odisha
2014 – 2016, secured 71.5%

CORE COMPETENCIES

- Data Pipeline Development
- Data Modeling
- Data Visualization
- Analytics Solutions Implementation
- ETL (Extract, Transform, Load)
- Data Migration
- Data Quality Assurance
- Data Visualization

IT SKILLS

- **Programming Languages:**
  Python, SQL, Pandas, NumPy
- **Software Tools:**
  Apache Spark, Apache Airflow, Databricks, Eclipse, Jupiter, Jira
- **Platforms:**
  Data studio, Data Lake, AWS S3, Git-Lab, confluence, Alation

SOFT SKILLS

- Analytical
- Team-oriented
- Multi-tasking
- Critical Thinking
- Problem-solving

Bangalore, Karnataka

sahumukeshkumar16@gmail.com

https://in.linkedin.com/in/mukesh-kumar-sahu

+91-9658143836

Profile Summary

- I have a total of 2.4 years of experience in Apache Spark, Apache Airflow, and Python libraries, including NumPy, Pandas, PySpark, SciPy, and Matplotlib, which I used for data wrangling and visualizations.
- hands-on experience on Jupyter, Databricks, Jira, Alation, and Google Suites.
- Skilled in optimizing data pipelines for effective data processing and analysis.
- Proficient in many approaches to codifying content, including oops, Data Structures and algorithms.
- In-depth knowledge of Database concepts, data lakes, big data, data management, and ETL processes.
- Gained a comprehensive understanding of the concepts of data science, advanced analytics, predictive modelling, and machine learning algorithms in multiple technical and functional domains.
- Conducted full lifecycle analysis, including data gathering and cleansing, deep-dive advanced statistical analysis and modelling, recommendations to optimize performance, and high-level design documents for the project.

Work Experience

Tata Consultancy Services, Bengaluru

Data Engineer / Developer	**Mar'2021-Jun'2023**

Key Result Areas:

- Developed and executed an end-to-end ETL process workflow (DAGS) using Apache-Airflow, Apache-Spark, Pandas, NumPy, and SQL.
- Conducting data processing and making the required adjustments enhanced the logic.
- Used GitLab to manage version control and CI/CD for each production release.
- Created generic code and modules that may be utilized across the project.
- Employed Pytest-Fixtures for generated validation scripts and unit test cases.
- Devised the workflow and schema for the intermediate data-cache and table, and offered full flexibility to the code by utilizing the dynamic and oops concepts.
- Liaised with coworkers and clients to elucidate the requirements for each task.
- Conceptualized and generated infrastructure that allows big data to be accessed and analyzed.

PROJECTS

Nielsen Tam Media (Annual UE), Bengaluru — Aug'2022-Jun'2023

Company: Tata consultancy services

Client: Nielsen

Description:

- An extension of the preceding project, where annual universe estimates were prepared for the United States market,
- The project has expanded the flexibility and sophistication of data analysis and set targets for sync data.
- Contributed to QA development by ensuring that all processes and data were more standardized.
- We utilized Jira-Board to track our progress while working in agile methodology and used Gitlab for CI and CD work.
- Architect schemas and datasets with the best way of partitioning to reduce the time of ingestion and execution by 60%.
- Given a new development to an existing module in which clients have their legacy code.
- Scheduled KT and code reviews with the operations team and client to create a generalized module and keep track of continuous integration to the current version of code.

Nielsen Tam Media (Local UE), Bengaluru — Mar'2021-Aug'2022

Company: Tata consultancy services

Client: Nielsen

Description:

- Worked on digital technology to devise and create a product that will automate Nielsen's weighting technique and prepare universe estimates for the United States markets.
- Building data pipelines (DAGs) through Apache Airflow with the help of Apache Spark, Python, Pandas, NumPy, and SQL
- I worked on data studio and data bricks on AWS, using Data Lake as a database with AWS S3.
- Contribute various scripts on Jupyter Notebook to perform data validation, data analysis.
- Build a generic module to perform Data Quality Assurance (DQA).
- Worked on Pytest and unittest for performing testing and Microsoft Excel for reporting.

CERTIFICATIONS

- Python for Data Science, IBM
- Data Analysis Using Python, IBM
- Data Visualization Using Python, IBM
- Python Project for Data Science, Coursera
- Applied Data Science with Python - Level 2, IBM
- AWS Cloud Quest: Cloud Practitioner, AWS